DOCUMENT RESUME

ED 224 833 TM 830 027

AUTHOR	Burstein, Leigh
TITLE	Using Multilevel Methods for Local School Improvement: A Beginning Conceptual Synthesis.
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	[82]
GRANT	NIE-G-80-0112
NOTE	52p.
PUB TYPE	Viewpoints (120) -- Guides - Non-Classroom Use (055)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Data Analysis; Data Collection; Elementary Secondary Education; Evaluation Methods; *Information Utilization; *Instructional Improvement; School Districts; School Effectiveness; Test Construction; Test Interpretation; *Test Results
IDENTIFIERS	*Unit of Analysis Problems

ABSTRACT

A conceptual synthesis is presented of the ways in which multilevel methods of data collection, analysis, interpretation and reporting can be used in instructional improvement activities in local school districts. This discussion of multilevel methods in school settings focuses on the "what" of information needs and usage. Principles and understandings regarding the analysis of multilevel data pertinent to information usage in local school improvement efforts are summarized. Prototypic educational contexts where multilevel methods might be used are described. Brief illustrations of usages of multilevel methods in local educational contexts are provided. Elaboration of the concepts and possible practices for applying multilevel methods in local school improvement efforts are suggested. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED224833

USING MULTILEVEL METHODS FOR LOCAL SCHOOL IMPROVEMENT:

A Beginning Conceptual Synthesis

Leigh Burstein

Grant Number

NIE-G-80-0112, P3

TM 830 027

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

CENTER FOR THE STUDY OF EVALUATION

Graduate School of Education

University of California, Los Angeles

CCI 2

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

This paper represents the first iteration of a conceptual synthesis of the ways in which multilevel methods of data collection, analysis, interpretation and reporting can be used in instructional improvement activities in local school districts. The basic premise for this effort is that while a variety of types of information (e.g., test, demographic and survey data) can be and is routinely collected from sources (e.g., pupils, teachers, administrators, parents) at multiple levels (individual, instructional groups, classes, school districts) of the local educational system, these data are not necessarily used in a manner consistent with existing knowledge about the possibilities and limits of data from multilevel social stuctures (see, e.g., Barr and Dreeben 1977, in press; Burstein, 1980a, 1980b, 1982; Roberts & Burstein, 1980).

Local school districts engaged in instructional improvement would appear to be settings particularly amenable to advances in practices and effectiveness through adopting a multilevel view toward the collection, analysis and interpretation of information on instructional context, programs and outcomes. Local educational personnel have the potential to obtain a detailed picture of the activities of the district which impinge upon its educational programs. Moreover, recent characterizations (e.g., Bank & Williams, 1980, 1981; Kennedy, 1982; Lyon et al, 1978) of the organizational responsibilities and actions of LEA evaluation activities and personnel certainly point to at least implicit knowledge of the linkages among information needs and usage at various levels of the educational system. That these personnel typically choose to better

represent the concerns of a subset of organizational entities (e.g., Superintendent and Board of Education concerns as opposed to local school-site decision-making) does not deny the po sibility of more comprehensive choice of clients and usage if viable and credible sources of information are available.

To facilitate the development of our conception, we suspend, for the time being, consideration of bureaucratic and organizational obstacles to comprehensive information use within school systems. These obstacles certainly exist and most likely will determine whether improvements in information use can be implemented. However, other investigators (e.g., Alkin, Bank & Williams, and Kennedy) are already carefully examining organizational operations and utilization and decision-making in shcool districts and there is little we can add to their insights at this stage. Their focus is on how school district evaluation activities operate and why they operate as they do. Our attempts at elucidating the potential of multilevel methods in school settings focuses instead on the what of information needs and usage.

The remainder of this report will proceed as follows. First, we briefly summarize certain principles and understandings regarding the analysis of multilevel data that may be pertinent to information usage in local school improvement efforts. Second, prototypic educational contexts wherein the multilevel methods might be used are described. Third, brief illustrations of usages of multilevel methods in local educational contexts are provided. The report concludes with suggestions for further elaboration of the concepts and possible

practices for applying mutilevel methods in local school improvement efforts.

Basic Themes in Multilevel Methods

In substantive area after substantive area, the recognition of the inherent interrelationships of individuals and social structure and the consequent commitment to a multilevel view of (and investigation of) social phenomena form the foundation for improvements in social inquiry (e.g., Cronbach, 1976; Merton & Lazarsfeld, 1950; Roberts & Burstein, 1980). The value of a multilevel perspective and multilevel analyses are especially evident in educational research and evaluation over the past decade. Advances in research on teaching and instruction (e.g., Burstein, 1980c; Corno, 1980; Cronbach, 1976; Cronbach & Snow, 1977; Cronbach & Webb, 1975; Greene, 1980; Martin, Veldman & Anderson 1980; Schmidt, 1980; Webb, 1980), and on schooling effects (e.g., Airasian & Madaus, 1980; Barr & Dreeben, 1977, in press; Bidwell & Kasarda, 1980; Brown & Saks, 1975, 1980a, b; Burstein, 1980a, b, c; Burstein, Fischer & Miller, 1980; Burstein, Linn & Capell, 1978; Burstein, Miller & Linn, 1981; Capell, 1981; Cooley, Bond & Mao, 1980; Cronbach, 1976; Keesling & Wiley, 1974; Madaus, Rakow & Kellaghan, 1980; Munck, 1981; Sirotnik, 1980; Wiley, 1970; Wisenbaker & Schmidt, 1978) are, in part, the results of a multilevel conception of how schooling works and the willingness to examine its elements and interconnections.

Multilevel considerations appear to be especially pertinent to the development of information-based methods for local school improvement. The extensive nesting and interactions among levels of the school system as well as the multiple sources (various participants, administrative records) and types (e.g., test, demographic, survey, interiew and observational data) of information obtainable from and aggregatable to the various levels represent both the obvious opportunity for informed decision-making, on the one hand, and confusion on the other. Under such circumstances, a sound understanding of the critical issues in the literature on data aggregation, cross-level inference, appropriate units of analysis and multilevel analysis are a useful starting point.

Recently, there have been a number of syntheses of relevant research that have focussed on the theoretical, conceptual and empirical bases for the inquiries into the impact of the multilevel character of educational systems on the measurement and identification of antecedents and correlates of educational performance (Airasian & Madaus, 1980; Barr & Dreeben, in press; Bidwell & Kasarda, 1980; Burstein, 1980b, 1980c; Cooley, Bond & Mao, 1981; Cronbach, 1976; Sirotnik, Nides & Engstrom, 1981). These authors build a case that the likelihood of understanding educational programs and their effects is enhanced by:

- developing better linkages between the conceptualization of an educational precess and its measurement (multilevel measurement of educational processes and contexts)

- utilizing an array of indicators of group (school, class, etc.) outcomes based on the notion that group mean indicators are potentially insensitive to differential outcomes associated with differential resource allocation strategies (alternative measures of group outcomes)
- employing test analyses procedures which are more likely to lead to instructionally sensitive outcome measures (multilevel approach to test development, analysis, interpretation and reporting)
- employing analytical models which can potentially identify effects at and within each level of the educational system (analytical models for multilevel educational data)

At this point, brief descriptions of selected empirical evidence about these four points is provided to further motivate our conception of systemic evaluation in local school improvement from a multilevel perspective.

Multilevel Measurement of Educational Processes and Contexts

The principle that the same variable can measure different constructs at different levels of aggregation is well-established (Burstein, 1980b, 1980c; Burstein, Fischer & Miller, 1980; Capell, 1979; Cline et al., 1974; Cronbach, 1976; Haney, 1974, 1977; Sirotnik, 1980). We have barely begun, however, to recognize the complexity of the task of measuring the desired construct (e.g., innovativeness, autonomy) at the desired level.

Both the problems in obtaining measures of the desired constructs from multilevel data and the benefits of a multilevel perspective have

been illustrated in several educational studies. Burstein, Fischer and Miller (1980) demonstrated the consequences for explaining pupil performance from distinguishing socioeconomic background (SES) as a measure (e.g., school aggregate SES) of the community context that conditions resources available for schooling from its use as a measure of potential home resources supportive of education (e.g., individual-level SES). By treating individual teacher responses as measures of their sense of personal efficacy and aggregate responses of teachers in mini-schools as indicators of the atmosphere and organizational structure of the mini-schools, Capell (1979, 1981) was able to detect patterns in the Alum Rock data (higher achievement associated with strong principal leadership, greater staff cohesion, less teacher autonomy, greater sense of personal efficacy (for teachers)) that were consistent with findings from other recent studies of the effects of organizational structure on student achievement (e.g., Kean, Summers, Raivetz & Farber, 1979).

Sirotnik and his associates (Sirotnik, 1980; Sirotnik, Nides & Engstrom 1981; Sirotnik & Oakes, 1981) found that it was necessary to take both the organizational (between-group) and psychological (within-group, individual) climate, as measured by scales completed by individuals, into consideration to account for differing school activities and quality under seemingly similar educational circumstances. Finally, Burstein (1980c) was able to point out a potentially serious misinterpretation in the evidence on the relationship between a student's typical degree of success (high, medium, low) while learning and academic achievement by distinguishing

success rates at the class level from the relative degree of success of students within classrooms.

The studies cited above are important for our present purposes in two respects. First, they demonstrate the value of multilevel methods in educational research and evaluation. Second, and more importantly, the measures used in these studies -- socioeconomic background indicators, survey responses from students, teachers and parents, classroom observation data -- represent typical information about educational processes and contexts that are or can be gathered in local educational settings. Apparently, they can be used to measure a variety of constructs -- home resources, community resources, organizational structure and atmosphere, personal efficacy, classroom and school climate, appropriateness of content, student level of functioning -- that are important in understanding and improving schooling when the linkage between the level of aggregation of measures and the construct of interest is clear.

The relevance of these concerns about the shift in variable meaning across levels is particularly pertinent to a variety of investigations of the effects of schooling that a number of local school districts have undertaken. Following up on interests from large-scale investigations of school effects (e.g., Coleman et al, 1966; Comber & Keeves 1973; Peaker 1975) and school effectiveness (e.g., Austin 1981; Brookover & Lezotte 1977; California State Department of Education, 1976; Edmonds, 1982; Klitgaard & Hall, 1973; Rasmussen 1974, 1976; Swanson, 1976), several school districts have conducted their own school effects (e.g., Kean, Summers, Raivetz &

Farber, 1979 (Philadelphia); Ramey, Hillman & Mathews, 1982 (Seattle)) and school effectiveness (e.g., White et al, 1975, 1976 (Atlanta); Gastright, 1977 (Cincinnati)) studies. The LEA-based studies seem no more (nor less) resistant to problems in measuring the variables of interest at the appropriate levels, and incorporating them properly in their analyses, than their large-scale, multi-site counterparts (see Burstein, 1980b, 1980c; Madaus, Airasian & Kellaghan, 1980; and Purkey & Smith, 1982 for discussions of methodological problems with these types of studies). If local educational agencies use these investigations to guide their decisions about instructional improvement programs and other school renewal activities, then one would hope that inattention to specific concerns about appropriate level of measurement and its relevance to construct-indicator match would have limited impact. Unfortunately, it is not clear that this will typically be the case.

Alternative Indicators of Group Outcomes

A number of investigators (e.g., Brown & Saks, 1975, 1980a; Burstein 1980a, b and c; Burstein, Miller & Linn 1981; Cline et al, 1974; Cooley & Lohnes 1976; Cronbach, 1976; Haney, 1974, 1977, 1980; Klitgaard, 1974; Lohnes, 1972; Spencer, 1992; Wiley, 1970) have pointed out that both questions of interest and technical considerations warrant analyses of aggregated data. The types of group-level indicators of processes and outcomes needed in any given study remains to be specified. In particular, when the purpose is to determine factors affecting pupil performance, analyses of between-group (class, school, etc.) means can hide important differences in the within-group distribution of pupil outcomes and

educational input. Groups receiving different treatments can have the same mean performance yet vary on other characteristics of the group's distributions (e.g., variance, proportion above or below a certain standard) or in the within relationships among educational inputs and outcomes. This is especially troubling given the ordinal properties of most metrics used to measure outcomes (See Spencer's (1982) discussion of technical problems with using means for comparisons of outcome differences).

One can again identify recent research relevant to the present investigation that demonstrates the importance of considering alternative indicators of group outcomes. For example, Brown & Saks (1975) used data on the level (mean) and variation (standard deviation) of a school district's performance on the Michigan Educational Assessment to demonstrate how variation in district policies may affect different indices of group outcomes. Brown & Saks (1980a, 1980b) have since elaborated their theory of production technologies and resource allocations within classrooms and schools in ways which highlight how the "tastes" of different educators cause them to maximize different outputs (e.g., to minimize the number of low achievers or ensure every student masters basic skills rather than maximize mean performance). The importance of considering a variety of indicators of group performance in the presence of divergent goals should be self-evident.

The use of alternative group-level indicators is also prominent in recent micro-analyses of resource allocations in classrooms and schools (e.g., Barr & Dreeben, in press; Bidwell & Kasarda, 1980;

Ferguson, 1980; Monck, 1981; Thomas, 1977) conducted by the Educational Finance and Productivity Center at the University of Chicago. These studies make the linkages of decisions about instructional grouping (both among classrooms and within classrooms in a given school) and pacing with the expected patterns of educational performance within schools explicit.

There are a number of situations that arise in school settings that suggest the hazards of automatic use of any single indicator to represent group context, processes and outcomes. For instance, Webb's (1980a, 1980b, 1981, 1982) detailed investigations of small-group interactions clearly demonstrate that groups with the same average levels but different ability compositions (e.g., a mixed ability group containing high, medium and low ability students contrasted with a uniformability) manifest different participation, interaction and performance patterns.

Another actual example derives from recent school effectiveness literature. In his studies, Edmonds (1982) focuses on differences between lower SES and higher SES students within a school in the proportion of students achieving mastery of designated educational objectives. Thus, if a substantial proportion of a designated group within a school performs at a prescribed level, the school is judged "effective." Moreover, despite his concentration on low income students, Edmonds fully expects middle income students to accrue even greater gains from the effective practices he prescribes.

Edmonds' interest in the antecedents of effective achievement of low-income students reflects a distinctive "taste" in the Brown &

Saks' sense. It led him to break a reasonably consistent practice of concentrating on school or district mean levels across all pupils in this body of research.

The point here is that although a concern for achievement may drive instructional improvement efforts, it is important to keep in mind the context for improvement efforts and be more refined in its objectives. Under most circumstances, a focus on raising the performance level of students around the middle of the overall performance distribution (e.g., say 40th - 60th percentile) will yield the highest gains in mean performance. Yet such a thrust ignores just the segment of students who have the greatest needs. A focus on the performance of the lowest quartile, on the other hand, devotes instructional resources in a manner likely to reduce the spread of performance (by establishing a performance floor or boosting more students over the minimal mastery point). Thus, the basic multilevel methods principle that group means do not account for all relevant group-level information should have a firm foothold in context-sensitive instructional improvement efforts.[1]

In our multilevel methods work, the proper handling and interpretation of scores aggregated over individuals has received most of our attention. Yet, in several respects, the logic holds as well for the content of instruction and of outcome measures. The choice between an emphasis on basic skills or on a broader array of knowledge has much in common with the decision about which group or level is of interest. Just as a focus on low income students at the school level dictates interest in certain indicators of performance (and perhaps

disinterest in others), judgments of the success of school improvement efforts can depend on the chosen level of aggregation over the content of instruction. It can also depend on the form of measurement of the content.

The concept of level of aggregation and measurement of instructional content can be envisioned as follows. Following a literal interpretation of the dictums of Title I, School A institutes a strong back-to-basics effort, adding more drill and practice activities (spelling quizzes, timed math drills, memorization and recitation of poems). The additional time School A devotes to these activities is obtained by foregoing most social studies, arts, science, and music instruction. School B, on the other hand, increases enrichment activities and attempts to enhance the breadth of its curriculum through dramatic play in its social studies work (e.g., various classes enact political campaigns prior to major elections, operate mock city governments, "live" through the experience of the Pilgrims, etc.). The teachers in School B tie in most lessons in reading, mathematics, and writing with these dramatic assignments but leave little time during school hours for drills on math, spelling and language arts facts.

Schools A and B make clear their instructional preferences. However, unless the array of non-teacher made tests (annual standardized tests, state assessment, district continuum) given in the schools are unusual, aggregate scores from these tests will differentially reflect the instructional emphasis and quality of the two schools. If two-thirds of a test's reading questions were devoted

to word identification, vocabulary and spelling and its math questions to basic facts and computations, total test scores might make it appear that School A is more effective because its students uniformly mastered their narrower curriculum material that predominated in total scores, while School B's students performed more inconsistently on their facts and mechanics. (Presumably, School B would do much better on more complex comprehension, reasoning and interpretation material that is covered by a more limited portion of the test.)

Unfortunately, this fictional scenario probably occurs all too often in current efforts to determine the content of instructionand its measurement in school improvement efforts. The premise here is that the level of refinement in distinctions about instructional content should be an explicit choice and the measurement of the consequences of instruction should be sufficiently refined to reflect both desired and unintended content distinctions. Aggregation over content in the scores from multidimensional tests is potentially hazardous if the intent is to determine what has been taught and how well. We return to this latter point in the next section.

There are other studies and other group-level indicators (e.g., Burstein, Miller & Linn, 1981; Cooley & Lohnes, 1976; Lohnes, 1972) that demonstrate the possibilities of improving explanations of the relationships among educational contexts, processes and outcomes. How to adapt at least the logic, and perhaps the methods, of alternative group-level indicators in local school program improvement activities clearly warrants further consideration.

Test Development, Analysis,Interpretation and Reporting.

The impetus for the work on multilevel approaches to test development, analysis, interpretation and reporting is the increasing concern about the instructional sensitivity of standardized achievement tests in instances mirroring the hypothetical scenario above. The weak evidence of schooling and program effects (Averch et al., 1972, Coleman et al., 1966; Stebbins et al., 1977) in the face of strong beliefs that students do learn from given school and program experiences has generated challenges to the instructional and program relevance of standardized achievement tests. Researchers knowledgeable about classroom practices and processes argue that as long as teachers have the freedom to choose areas of coverage and emphasis, tests cannot be expected to have relevance for all classrooms. For similar reasons curriculum developers conclude that tests are not appropriate to the content of their curricula. These arguments have intrinsic merit but raise as many questions about the appropriateness of instructional coverage decision of teachers and curriculum developers as they do about the utility of the tests for measuring skills that should be part of the repertoire of the nation's students.

Another complaint about instructional sensitivity of standardized achievement tests focusses on the psychometric basis of their construction (Carver, 1974, 1975; Cronbach, 1976). Carver (1974, 1975) asserts that these tests are constructed by traditional psychometric methods to be sensitive to individual differences and as a result, are unlikely to be sensitive to the instructional and program differences across classrooms and schools. Recent work by

Airasian, Madaus and their colleagues (e.g., Airasian & Madaus, 1976, 1980; Madaus, Rakow & Kellaghan, 1980; Rakow, Airasian & Madaus, 1978) support Carver's contention in that they found clear evidence of differential instructional impact of schooling when between-class and between-school variation in test performance was separated from individual variation.

The potential benefits of multilevel approaches in understanding and describing test peformance becomes more pronounced when investigators move away from test scores that are horizontal aggregates of many subskills and competencies (e.g., sums over subtests) as we have urged in the previous section. Multilevel examinations of test item data have the potential to lead to better informed test development, analysis, interpretation and reporting procedures. For example, careful investigations of test item and subtest data appear to e particularly valuable for detecting effects due to background differences (e.g., prior learning, sex, socioeconomic and demographic differences), instructional coverage and emphasis, and instructional organization (e.g., grouping and pacing effects). When these separate effects can be identified, it may be possible to construct composites of items which are potentially sensitive to the context factors of instruction and describe performance accordingly. Likewise, test developers (either test publishers or LEA and SEA personnel) could include procedures that guard against unknowingly selecting items influenced by "irrelevant" context and situational characteristics (where "irrelevancy" is determined by the purposes for which the test would be used). At the least, developers would be better able to describe the properties of

their tests after carrying out a multilevel examination of their properties.

The most promising avenues of work on multilevel approaches to testing appear to be capable of combining the features of less technical empirical descriptions of test performance (Harnisch & Linn, 1981; Miller & Burstein, 1979; Miller, 1981) with recent advances in information processing models for test design and interpretation (Glaser, 1981; Brown & Burton, 1978; Tatsuoka & Birenbaum, 1980; Baker, Linn & Quellmalz, 1980). For example, Miller (1981) found that an examination of the class-level and individual-level patterns of correct and error responses by 5th graders on the items of a fractions test enabled him to identify classrooms that followed an atypical instructional sequence (e.g., heavy emphasis on work on solving for unknowns in fraction equations like $5/n = 2/3$) and classrooms where instruction on addition of fractions was either faulty or lacking (e.g., most students in a class added in both the numerator and the denominator).

What was particularly interesting about Miller's investigation was that the differences among classrooms were evident in graphical representations of the patterns of test item responses modeled after techniques developed by Japanese psychometricians (Harnisch & Linn, 1981; Sato, 1978, 1981; Tatsuoka, 1978) for use by teachers. To be sure, statistical indicators such as class-level analogues of Sato's caution index and intraclass correlations were used in the study, but these statistics merely mirrored the graphical results. It appears to be possible to adapt these methods for local program improvement in a

way that is easily understood by school building personnel (teachers, principals).

We see no need to recount separately the rationale for and developments in analytical models for multilevel educational data. The preceding discussions and much of the literature delineate the need for improved methodology and its consequences. The substantive investigations described in earlier sections typically combined a better conceptualization of the multilevel character of educational data with better analytical procedures to derive their contributions to existing knowledge. In virtually every case, the combination involves adapting analytical machinery to the substantive educational phenomena under investigation rather than the opposite paradigm typical in earlier research.

Prototypic Educational Contexts

At this point we radically shift directions and delineate prototypes of the available information and its presumed constituents in school systems. This section draws heavily on materials obtained from school districts participating in Bank & Williams' Organizational Study (Bank & Williams, 1980, 1981), from Kennedy's (1982; Kennedy, Aaling & Newman, 1980) studies of evaluation and testing activities in school districts and from a high school in a local suburban school district as part of a beginning CSE-UCLA Partnership collaboration.

Informational Elements

Certainly, school districts, and hence schools, vary in the types of information available for decision-making at various levels (class,

school, district, etc.). But under the most typical circumstances, one can expect a school district to collect the following:

1. Student demographics - age, sex, ethnicity, home language, parental occupations, eligibility for AFDC reduced price lunches, medical histories; in some districts, home address (in some cases parental education and family size).
2. Teacher and building-level administrator background - age, education, previous employment and educational history, special certification (bilingual) and subject-matter expertise.
3. School building characteristics - information about physical plant (e.g., age, capacity, particular resources), student body and community composition (ethnic composition, class sizes, neighborhood wealth, community involvement in neighborhood schools).
4. Program participation - availability, staffing and funding of special programs (e.g., Chapter I, Title IV-C, Bilinguil Education, Degregation funding) at local school site.
5. Standardized norm-referenced tests - typically given either Fall and Spring or Spring only in several grades in primary (e.g. 3 and 6) and secondary (e.g., 8, 10 and 12) school years. Scores reported at individual, class and school levels.
6. Criterion referenced testing - typically district derived and tied to objectives from the district's curriculum continuum. These testing programs generally are implemented only in primary schools.
7. Minimum competency and proficiency testing - conducted in secondary schools primarily (grades 8, 10 and 12).

8. Group and individual ability or aptitude testing - done to determine pupil eligibility for special programs (e.g. Special Education or Gifted Programs in Primary Grades) and placement decisions (e.g., College Admissions and track placements in secondary years).
9. Teacher-made tests and curriculum embedded tests - strictly available within classes in most schools.
10. Curriculum information - curricular packages and texts used in classrooms, topic coverage from continuum (assumed or measured).
11. Course taking patterns - available from student cumulative folders and from prescribed offerings.
12. Grading practices - teacher reports of student grades.
13. Class materials and resources inventories - e.g., audio-visual equipment available in classrooms and schools.
14. Student measures of classroom and school climate - student responses to surveys about class and school environments and other aspects of their educational experience.
15. Teacher measures of classroom and school climate and activities.
16. School building administrator measures of school climate and activities.
17. Parental surveys of perceptions and support of school activities.
18. Parental participation in school activities.
19. Administrator observations and evaluation of teachers.
20. Teachers' observations of other teachers.
21. District personnel's observations and interviews of building personnel.
22. Surveys of graduates to determine occupational and educational status.

Obviously, school systems do not uniformly maintain all of the above sources of information. It is even less likely that much of the information is regularly used for the multiple purposes for which it may be relevant. Studies by Kennedy (1982) and Lyon et al (1978) suggest that usage patterns for all but the routine standardized test data and other information compiled to fulfill evaluation requirements for federal and state categorical progams is sporadic and unsystematic at best.

Exemplary Existing Information Systems

Nonetheless, there are districts that do attempt to maintain conprehensive information systems that appear to be used routinely at multiple levels of the school system. For example, Clark County School District, Las Vegas, Nevada and the Cincinnati Public School System operate such information systems.

The Clark County System routinely collects and uses the information types indicated in Table 1. In addition, the Elementary Education Division developed a comprehensive management and accountability system for elementary school teacher and principals and central office administrators. This system is operated under the general guidelines of a district-determined "Elements of Quality for Elementary Schools" that permeates the systematic instructional management objectives and the comprehensive handbook that illustrates and describes its major components (Clark County School District, 1976).

While the information system is more developed in the elementary schools in Clark County, analogous management information systems are

TABLE 1. Information routinely collected
by the Clark County School District, Las Vegas, Nevada

I. Achievement Data

A. NRT's

1. Otis - Lennon School Ability Test (OLSAT) - Grades 2,3,5
2. Otis - Lennon Mental Ability Test (OLMAT) - Grades 8,11
3. Stanford Achievement Test (SAT) - Grades 3,6
4. California Achievement test (CAT) - Grades 8,11

B. CRT's

1. Reading - Elementary - Grades 2-6
2. Math - Elementary - Grades 2-6
3. Language Arts - Grades 4,7
4. Nevada State Proficiency Examination - High School
5. Optional Computer-Assisted Test Construction

C. Management Information Systems - Placement & Diagnostic

1. Reading
2. Math
3. Language Arts - Scope & Sequence; pre- & post-tests - Grades 7, 8

D. Title I Reading Student Progress Report

E. English Support System Profile Sheet - Secondary

F. School Roster Report - Nevada High School Proficiency Exam

G. Kindergarten Progress Report (report card)

II. Non-Achievement Data

A. Performance Evaluation Report - Elementary School Administrator
B. Criteria for the Assessment of Instruction Checklist
C. Principal's Supervisory Log (goes with B.)
D. Plan to Achieve a High Priority Objective
E. Teacher Performance Criteria Log - Secondary
F. Administrative Performance Criteria Log - Secondary
G. Principal's Observation Sheet
H. Teacher - School Profile
I. Guide to Supervision & Evaluation of Certified Employees
J. Report of Teacher Personnel Records Audit
K. Parent Opinion Survey
L. Parent - Teacher Conference Report
M. Teacher Opinon Survey

taking hold at the secondary level as well. A particularly noteworthy component was the Syllabus-Based Testing system developed for English, General Mathematics, Algebra, Biology and U.S. History in the ninth and tenth grades. Teacher task forces designed a descriptive syllabus for each course and a computerized test item bank which related items specifically to stated curriculum objectives was developed. Each teacher administering a particular curriculum test was allowed to omit from the test any material not covered in instruction at the time of administration. The diagnostic value of this procedure is that the school district can learn how well students do in areas in which they receive instruction and also determine variation in patterns (timing and sequence) of instructional coverage. While the syllabus testing is too new to determine its value (both in context and for transfer to other settings), it does reflect a sensible multilevel and multi-purpose usage of information generated from the district's competency-based educational management system.

Cincinnati's system for information management and use has a longer history than Clark County's and is, in many respects, at least as elaborate and perhaps more multifaceted. The unique features of the Cincinnati system are their decentralization of the operation of information usage activities to the local school level, the creation of the position of the Local School Evaluator (LSE) to serve the evaluation needs of individual schools and, recently, the formation of Local School Budgeting (LSB) to decentralize the process of educational planning while at the same time involving each school's total community.

Felix (1982) summarizes that rationale and the activities of LSE and LSB and the ways in which these entities call upon the information resources of the district (including the Cincinnati Instructional Management System reports, Chapter I School Profiles, special item analyses and reporting, annual surveys of administrators, teachers, parents and students). The information sources petinent to a particular school (including a variety of demographic indices (e.g., dropout rates by cause, socio-economic variables) are combined to form the district's School Information System. The school receives a comprehensive report of its data each yer and are helped by the LSE to interpret and apply its information. Clearly, a sense of the utility and purpose of credible and appropriate information is conveyed by Cincinnati's activities.

The User Community

The discussion thus far has focussed on the possible elements of information systems, i.e., what can be included in systems created for use in instructional improvement. Little has been said about who would use available information and how. But certainly these questions are central to a decision to implement such a system.

District Level. Various investigations of evaluation activities in school districts (e.g., Bank & Williams, 1980, 1981; Kennedy, 1982; Lyon et al., 1978) have tended to address the user question almost exclusively from the perspective of the central administration personnel (Superintendent, Curriculum departments, Research and evaluation units) at the district level. The body of accumulated conventional wisdom regarding district-level (and beyond, e.g., State

and Federal) uses of information is quite extensive. While some pessimism about the use of information for decision-making at the district level persists, the evidence from the cited studies is that most districts establish organizational and informational structures that tend to be aligned with routine information needs of district-based decision-makers (e.g., School Boards, Superintendents, local media).

School Level. Unfortunately, less is known about the actual needs and usage of available information at the school level. More specifically, with the exception of vaious studies of teacher uses of tests (e.g., Dorr-Bremme, 1982) and incidental commentary from broader studies of schooling (e.g. Goodlad, 1982), there is hardly any empirical evidence about how the teachers, administrators, other staff and parents do use available information resources. Consequently, little can be said at this point about how these groups could use credible and timely information. Yet, the norm in American education clearly points to the school (and perhaps the classroom) as the central organizational entity in translating educational and instructional policy into practice. One way of characterizing the available wisdom on educational change is that a culture that inculcates instructional improvement principles must operate at the local school site for constructive programmatic change to occur (Berman & McLaughlin, 1981).

If the individual school is the functional locus of instructional improvement efforts, then a viable information system must at some point become responsive to the needs at this level of the school system. The information system may be viewed as a club or a "goad" by

school site personnel, as many view competency and proficiency testing programs. Alternatively, it may be treated as a resource to help the school community understand where it is, where it could be and how it might get there. The point is that while information systems can be perceived to be benign or malignant, and this perception may affect internal resistance to usage, it is nonetheless functional if it can credibly guide the school in valued directions.

Elementary/Secondary Distinctions. One can certainly cite a variety of ways in which school level personnel might make use of a sufficiently comprehensive information system (Cincinnati already illustrates how such systems might operate). We will provide a few examples later on. To enhance a general conception for the foundation of such systems, it is useful to distinguish between elementary and secondary schools in terms of uses and users of information.

At the risk of seriously understating their diversity, elementary schools typically operate organizationally according to a unidimensional hierarchical structure. A principal (and perhaps a special programs coordinator) provides general direction to the school's instructional activities and monitors specific schooling practices only to the extent that they comply with district, state and federal program guidelines. While a district might mandate a prescribed curriculum through their continua, it is rarely the case that the principal imposes his/her own curriculum on teachers. As long as a teacher's actions and instructional decisions fall within broad bounds of educational practices, the principal is unlikely to intervene substantially in the day-to-day activities of individual classrooms. The principal's function is clearly to lead and manage.

Under the typical organization of instructional activities in the elementary grades, the teacher controls virtually all of the in-school educational experiences of the students in his/her classroom. In the early grades, teachers tend to devote a substantial amount of time to monitoring and easing the intellectual, behavioral and emotional development of pupils (in the child development sense), socializing each child to the student role, and preparing them for the future accumulation of complex knowledge. Since the pupils and teachers tend to stay together throughout every school day, the teacher must work to develop the social system within the classroom so that all students can learn participation and cooperation skills. And, by the way, under the best of circumstances, these same teachers are subject-matter specialists as well.

The instructional and child development process in elementary grades are dominated by the organization and activities within the classroom.[2] The teacher's role in this setting has been characterized as both that of a clinical problem solver (Berliner 1975; Shavelson, 1976) and of an executive (Berliner, 1982). Under either characterization, the teacher is viewed as an information synthesizer who must combine various bits of empirical data with her/his own insights to arrive at both short-term decisions and long range plans.

While their classroom responsibilities may be more diverse, current patterns of information usages by elementary school teachers focus almost exclusively on test data. That is, the variety of non-test sources delineated in previous sections rarely impinge on the day-to-day functioning of teachers in classrooms. Their use of test data beyond curriculum embedded and teacher made tests is also rather

specialized (e.g., Dorr-Bremme, 1982). Standardized test data is used primarily for pre-year planning and initial grouping of students, if at all.

According to Dorr-Bremme (1982), testing programs are useful in the classroom to the extend that they take into account teachers' routine thinking and practices in assessing students' achievement. He concludes that such testing activities would need to be

(1) proximal to the everyday instructional tasks teachers need to accomplish

(2) consonant from teachers' perspectives with the curriculum that teachers are actually teaching

(3) immediately accessible to teachers

(4) designed to include a variety of performance "contexts"

Most present standardized testing efforts fail to meet these needs either through lack of timeliness or relevance (Airasian & Madaus, 1983; Seewald, 1981; Schmidt, 1983). Obviously any information systems intended for classroom use that incorporates testing program data will need to be sensitive to these concerns.

To summarize, the current organization and operation of elementary schools places most instructional and child development decisions under the control of the classroom teachers. These teachers' uses of information in their management and decision making activities relies heavily on those sources that can be tied into their routine practice in monitoring student achievement. Unfortunately few of the information sources accessible to school systems fit this purpose. Thus the development of comprehensive information systems relevant for instructional improvement in elementary school settings must do one or all of the following:

(1) Reduce the dominance of the classroom teachers in everyday instructional decisions through having (a) school-level administration play a more active instructional leadership role, (b) greater coordination among teachers within and across grades on instructional planning (content, sequence, etc.), and (c) more explicit curricular and educational policies to which individual teacher must be ultimately responsive.

(2) Increase school building personnel's (teachers, administrators) understanding and knowledge about various types of information and how they might be used in instructional improvement.

(3) Improve the timeliness and relevance of information systems to the specific content in the individual school setting.

Persons familiar with recent literature on educational change (e.g., Berman & McLaughlin, 1981; Goodlad, 1982; Sirotnik & Oakes, 1982), school effectiveness (Edmonds, 1982; Purkey & Smith, 1982) and linking testing with instruction (Airasian & Madaus, 1983; Burstein & Miller, 1982; Leinhardt, 1982; Leinhardt & Seewald, 1981; Schmidt, 1983) should recognize the connections between the above points and the concerns highlighted by these lines of inquiry.

The organizational structure and division of labor in secondary school settings bring a different combination of influential parties into the educational development and change processes. Relevant school personnel tend to fall into three distinct groups. First principals and other administrators become even more management-oriented although the array of activities they must coordinate is much more diverse. They are even less likely to be instructional leaders as subject-matter specialization and non-academic, extra-curricular involvement of students increases.

In Secondary schools, the teachers role shifts from instructional and environmental manager to that of curriculum specialist. Teachers

some into contact with a greater number of students for a shorter period of time, perhaps spread over a greater number of years. Concerns about the development of the students play a back seat to the transmission of knowledge and skills defined as pertinent to the specialized curriculum interest. The disciplines (mathematics, science, English, history, etc.) dominate teacher thinking and actions. Teachers transmit content which students are expected to absorb and to demonstrate their mastery through class examinations and other assignments. These examinations and assignments are directly linked to grading with little concern for diagnosis and instructional prescription and limited remediation offered. Indirectly, the teachers through their instructional and grading practices, provide the foundation for the sorting process that assumes an increasingly prominent role in later high school years.

The third group of potential information users in secondary schools are the counseling and guidance personnel. Where the elementary schools may operate a number of specialized programs (e.g. compensatory education, special education, bilingual education) that provide other influential people in the educational development of specific targeted students, the counseling and guidance functions in secondary schools ideally serve the needs of the entire student body. Counselors participate in developing students' courses of study, monitoring their participation in school activities, their socialization and social behavior (e.g., truancy, vandalism, disruptiveness), counseling regareding persopnal problems, and aiding in the determination of future educational and career plans. In essence they assume all the responsibilities for the social and educational development, besides curriculum content, which are

handled by teaching personnel in elementary schools. They were also more likely to conclude that tests motivated students to study harder, the pressure of testing has beneficial effects and tests of minimum competency/proficiency/functional literacy should be required of all students for promotion or high school graduation. Clearly, these usage patterns and attitudes are reflections of the departmentalization in secondary schools with attendant perceptions of subject matter expertise and interest and reduction in a sense of responsibility for students' educational and social development.

Counselors, on the other hand, tend to be more interested in a broad array of indicators of individual students. They, more often than teachers, find themselves in the role of interpreting empirical information (test scores, interest inventory results, career and college choices) with individual students. They may also find themselves in the position of accumulating and interpreting information aggregated over relevant subgroups of students for other instructional and administrative personnel. For example, counselors are more likely than teachers to monitor trends in test performance, truancy, participation in extra-curricular activities, and career and college choices.

The information use by the principal in secondary schools is less easily typified. Certainly they are aware of test performance patterns, student activity patterns, attendance and incidence of disruptive and destructive behavior in the school. It appears, however, that while their direct participation in the ongoing activities, either instructional or extra-curricular, is likely to be limited, their role in the establishment of environment for schooling

(as a workplace, as a supportive or accommodating locus for adolescent development along with the development of intellectual and career interests) is probably quite pronounced. One would think that the need for the kinds of comprehensive information we envision would be substantial for the administrative leaders in secondary school settings but evidence at this level is sorely lacking.

Our characterization of the prototypic organizational structure, division of responsibilities and activities among personnel and their typical uses of information was intended to support a view that comprehensive systems involving multilevel data have a role in instructional improvement efforts. More importantly, the potential distinctiveness of the socio-organizational mechanism and information needs in elementary and secondary settings was highlighted. Thus, we should approach any endeavor at implementing informational systems for instructional improvement with a clear understanding of the explicit contextual distinctiveness between these levels.

Multilevel Methods in Local Educational Contexts

At various points in our discussions of themes in multilevel methods, we pointed to actual incidents where these methods have been employed in actual educational settings. Although most of the examples depicted usage at levels above the individual school, these were exceptions. For instance, schools operating according to Edmonds' (1982) definitions of school effectiveness use indicators other than school means in their decision-making, as did the principal described in Footnote 1.

Existing Deterrents

The other two general categories of multilevel methods we discussed; namely, multilevel measurement of educational processes and test development, interpretation and reporting, are seldom considered in these settings prmarily because the relevant information is either not collected or retained in a form suitable for usage at the school levels. For instance, even in those school settings that routinely gather survey information about teacher, parent and student attitudes, rarely feed this information back to school buildings in a form conducive to usage in the manner described in the multilevel measurement discussion. To be informative for the school, these data would need to be reported in a variety of ways that reflect the normative climate of classrooms and the school as a whole (e.g., the levels and homogeneity of satisfaction of students, teachers, parents and administrators with the goals, direction and atmosphere in the classroom and in the school) relative to the existing patterns of such indicators for classes and schools operating under similar socio-demographic circumstances. They must be provided to all relevant parties (teachers, administrators, parents, students) at the secondary level who would be expected to participate in educational change efforts.

The problem leading to the lack of application of multilevel methods in test development, interpretation and reporting stem from a lack of timeliness and relevance. Moreover, the generally inadequate knowledge among school-based educational personnel regarding educational testing principles and appropriate practices could

severely limit the usage of even high-quality test information that is both timely and relevant. Thus, even if the test analysis and support services available in Japanese schools (as described by Sato) were available in American schools, a staff development effort on testing for educational decision-making that rivals present attempts to achieve computer-literate instructional personnel might be necessary.

Clearly, the deterrents to improved information use are substantial. And, since the principles of multilevel methods manage to escape the understanding of even the research community, it will not be easy to bring about their implementation in local school settings which have only limited and narrowly focused information usage patterns.

Nonetheless, the current underutilization of available information in instructional improvement and educational change may be a blessing. Since there is limited conventional wisdom beyond that associated with various testing activities, it will not be necessary to extinguish many poor practices and thus have to battle entrenched interest groups about the substance and operation of information systems. Moreover, these systems will be introduced in settings seeking to improve and that are thus conducive to changes in operation and functioning.

We are not yet intimitably acquainted with any educational setting that has already undergone or is well along in their instructional improvement and change activities, nor are we aware of districts where these processes are in operation that have information systems targeted to improvements at the building level[3]. We have

begun work with a high school (Site A) where there is an interest in educational change and instructional improvement, and perhaps the potential resources for a comprehensive information system to serve as part of the improvement effort. In the remainder of the paper, we briefly describe Site A and its potential for integrating multilevel information usage into the change process.

Site A

Demographic Conditions

Site A is one of two high schools in a suburban school district. Approximately 1800 students are enrolled in Site A of whom approximately 85 percent are Anglos. The community is predominantly lower middle class (upper blue collar, lower white collar) at present. School authorities estimate that 15 to 20 percent of the graduating seniors go on to four-year universities and colleges while approximately 20 to 30 percent attend a local community college. The community has no heavy industry but does contain a few companies involved in the manufacturing and marketing aspects of high technology business.

The student body in Site A is quite diverse. The school prides itself on offering a full range of advanced placement courses (sciences, mathematics, foreign languages, English, history) as well as its successful athletic programs and other extra-curricular participation. Nonetheless, there is at least some latent concern in the district and the school that Site A graduates may not be pursuing

post-secondary educational and career opportunities that reflect the quality of the school programs. There is also some unease over the fact that of a cohort of 600 who start at Site A in the 10th grade, only about 350 remain as part of the graduating class at the 12th grade.

School Resources and Special Programs

One distinctive feature of the resources at Site A is the presence of a computerized individual information system. CASA (Computerized Accountability for Student Achievement) is a computer-based student information management system which provides counseling staff and assistant principals instantaneous access to a wealth of information stored on each student. Through this system, counseling staff and administrators can access such student-level information as standardized test scores for multiple years including information prior to entry to high school; proficiency testing information (from the District's Computer Managed Instruction program, which monitors student progress in Grades K-10 through the use of CRT's based on District continuum); curriculum and performance information including courses taken, credits, grades and class rank; background information including parental occupations, family size, census tract location of residence, ethnicity; current school status information including eligibility for special programs (Gifted and Talented, Bilingual, Special education) special school activities (athletics, school paper, etc.), complete attendance history and referrals to various school services (psychologist, health office counselor and guidance office, principal, etc.).

The CASA system was developed with Title IV-C funds. The system is currently being augmented through state school improvement funds

targeted to the development of a computer-managed Career Magnet School (CMS). CMS seeks to align student career interests with competencies required by specific careers and, in turn, align these competencies with specific courses at Site A in which they may be obtained. Over the past year, teams of teachers have been developing the lists of skills associated with given courses in the school and an analogous list required for each career program (Careers Magnet schools are first broken down into career paths such as computer technology and then further into programs such as computer technician, operator, programmer, or designer). The skill-to-course and skill-to-career matches are being computerized so that counseling staff can develop courses of study for students choosing particular careers and monitor their progress at obtaining prerequisite skills.

CMS is one of two efforts recently instituted at Site A that are directed at career and academic decision making. The other, the Learning Resources Center (LRC) offers students and teachers academic resources (materials, assistance) for remedial and advanced work. These two programs appear to be guided by a general concern that students do not have the necessary information and skills to attain post-secondary education and careers they might want. The implicit assumption is that providing students more information about career opportunities and their necessary prerequisites (CMS) on the one hand, and resources for remediating or enhancing their academic performance (LRC), on the other, that students will make better decisions about how to benefit from their high school experience and be better prepared for their future.

Staff Concerns

Most of the Site A activities recounted to this point were primarily created for and used by the counseling and administrative staff. Teachers had little role in the creation of CASA and CMS other than their participation in the specification of skills related to courses and careers and of proficiency tests used by the district. Moreover, though there is strong interest in providing both computer literacy and programming training to students, the instructional resources for these efforts are limited, especially when contrasted with those available to the school's counseling staff.

Discussion with departmental chairs suggest that teachers may have certain concerns that seem to be manifested by the same sorts of problems that perhaps underlie the development of the CMS and LRC programs. After a series of meetings, the chairs settled on three topics of concern that might be dealt with in collaboration with UCLA personnel[4]:

1. Test Evaluation -- How can the test information gathered from students be used more appropriately for various forms of instructional decision-making? This concern focusses around teacher-made tests but the counseling staff raised similar questions about the test information contained in their computerized information system that might aid in the career counseling process.

2. Student Variability -- How can the school deal with the variability of the backgrounds of students in its course offerings and instructional emphasis?

3. Critical Thinking Skills -- How can the school develop ways to build critical thinking skills involving reading and writing across subject matter areas?

These questions certainly reflect the academic focus of secondary teachers. However, they are clearly symptomatic of the teacher's role in addressing the broader questions that both guided the counseling staff and led the school to seek collaboration with UCLA. In other words, the various educational constituencies at Site A seem to want to address, from their own perspective and within their own repertoire of interests and skills, the following:

> Are students being provided the kinds of educational experiences that will enable them to successfully pursue their career goals?
> Are community needs for an educated citizenry being met?

Information Context and Needs

The issues that guide Site A's school improvement efforts would appear to be the kind that would benefit from better information use. Moreover, with a few notable exceptions, the kinds of information that improvements require is already being gathered and, as a result of CASA and CMS, is readily accessible at least in some form. The exceptions are the absence of the kinds of attitude and opinion - surveys of students, teachers, administrators and parents that might provide a better indication of the climate of opinion in operation in the school and the community, determine whether the planned courses of action adequately reflect perceived needs and monitor the progress of the change efforts. Surveys of both graduating and entering students

would seem to be especially valuable given the substantial focus on providing students better guidance for career and academic decision-making.

Once shortcomings in the gathering of survey data are resolved, it should be possible to develop a comprehensive information system that serves the interests of all educational constituencies at Site A. The system would certainly need to provide information relevant to the issues identified. Moreover, since these issues are probably typical of those that interest other high schools (with the exception of the apparently limited interest in violence, vandalism and substance abuse), the broad parameters of the system should be generalizable to other settings.

Concluding Comments

We started out to develop a conceptual synthesis of the ways in which multilevel methods might be used in local school improvement efforts. Along the way we have expanded the analytical reasoning regarding the conditions under which a multilevel view of educational data might lead to better information use in the study of schooling. Certainly the pertinence of multilevel methods to the understanding of the effects of education in becoming a more commonly held perception across a wide array of educational professionals (researchers, policy makers, and practitiners).

Our exercise in enlightenment bogged down once we shifted interest from studying and understanding schooling to its practices. As characterized in our descriptions of prototypic educational

contexts, school cultures involve a variety of sources of information, and a number of pertinent constituencies with clearly demarcated responsibilities that might be able to use properly collected and routinely accessible information. But little is known about how school building personnel operate in a context with high-quality, timely, pertinent information either because these conditions don't exist or no one has yet documented how school-level personnel respond under ideal information conditions. Perhaps the next stage of our work on systemic evaluation will begin to remediate the state of ignorance on these matters.

FOOTNOTES

[1]It should be clear that one does not have to be devoted to compensating for the inadequate performance of low income pupils to derive benefits from interest in the distribution of performance rather than simply its level. In their annual reports, the California Asessment Program provides schools with the quartile distributions of the performance of 3rd grade students along with a variety of mean indicators (overall and for various demographic subgroups). A former principal of a suburban, typically higher performing school pointed out that while the school's overall performance each year (typically above 70 per cent correct, which is above the 90th percentile statewide) didn't tell him anything, he did keep track of the number of children that fell in the lower quartile each year because this meant that there were still students who needed to improve. It should be clear that even high achieving schools can benefit from an awareness of the functioning of their weakest students and school mean performance doesn't typically capture this type of information.

[2]Our emphasis on the classroom as the socializing environment is not intended to denigrate the role of the home although school hours represent a significant portion of the waking hours of school-age children. Moreover, once television time and play time with peers are excluded, the active participation of the parents during these elementary yars is diminished even further.

The school's influence on socialization, in contrast with the classroom's, is largely indirect. Different school have different emphases on cross-classroom and cross-grade activities (although it is a rare class whose students remain segregated during recess and lunch periods). The school's impact on educational and social development is also felt through the mechanisms it institutes for the transitions from grade to grade. To the degree that the atmosphere is one in which class formation policies are centrally coordinated and teachers actively coordinate educational decisions across grades, the school can play a strong role in development over time.

[3]The Cincinnati school activities described earlier might fit this description. It is also possible that some of the school systems identified by Edmonds (e.g., New York, Milwaukee, St. Louis, New Haven) that are participating in the effective schooling efforts might also be operating in the manner we envision. We simply don't have enough information about them.

[4]There are potentially multiple avenues of cooperation and collaboration between Site A and UCLA. Site A is located in a district which is a member of The Partnership, a consortium of school districts, community colleges, county offices of education and the Laboratory of Community and Schooling in Education, a unit of the UCLA Graduate School of Education. CSE and Laboratory staff may both be collaborating with Site A.

REFERENCES

Airasian, P. W., & Madaus, G. F. A study of the sensitivity of school program effectiveness measures. Report submitted to the Carnegie Corporation, New York. Chestnut Hill, MA: Boston College, School of Education, 1976.

Airaisian, P. W., & Madaus, G. F. 1983

Austin, (1981)

Averch, H., Carroll, S. J., Donaldson, T., Kiesling, H. J., & Pincus, J. How effective is schooling? A critical review and synthesis of research findings (R-9560PCSF/RC). Santa Monica, CA: The Rand Corporation, 1972.

Baker, E., Linn, R. L., & Quellmalz, E. Knowledge synthesis: Criterion-referenced measurement. Center for the Study of Evaluation, University of California, Los Angeles, CA 1980.

Bank, A., & Williams, R. C. Evaluation design project: School district organization study. Los Angeles, CA: Center for the Study of Evaluation, University of California, 1980

Bank, A., & Williams, R. C. Annual report - Evaluation design: Organizational study. Center for the Study of Evaluation, University of California, Los Angeles, 1981.

Barr, R., & Dreeben, R. Instruction in classrooms. In L. S. Shulman (Ed.), Review of Research in Education, Vol, 5, Itasca, IL: F.E. Peacock Publishers, Inc., 1977, 89-162.

Barr, R., & Dreeben, R. (in press)

Berliner, D. C. The beginning teacher evaluation study: Overview and selected findings 1974-75. San Francisco, CA. Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1975.

Berliner, D. C. The executive functions of teaching. Paper presented at the National Invitational conference on translating Research on Teaching into Practice, Arlington, VA, 1982.

Berman & McLaughlin 1981

Bidwell, C. W., & Kasarda, J. D. 1980

Bidwell, C. W., & Kasarda, J. D. 1981

Brookover, W. B., & Lezotte, L. W. Changes in school characteristics coincident with changes in student achievement. East Lansing: Michigan State University, College of Urban Development, 1977.

Brown, B. W., & Saks, D. H. The production and distribution of cognitive skills within schools. Journal of Political Economy, 1975, 83, 571-593.

Brown, B. W., & Saks, D. H. Production technologies and resource allocations within classrooms and schools: Theory and measurement. In R. Dreeben & J. A. Thomas (Eds.), The analysis of educational productivity, Vol. I: Issues in microanalysis. Cambridge, MA: Ballinger, 1980a.

Brown, B. W., & Saks, D. H.. Economic grouping and data disaggregation. In K. H. Roberts & L. Burstein (Eds.), Issues in aggregation, No. 6, New directions for methodology of social and behavioral science, San Francisco: Jossey-Bass, Inc., 1980b.

Brown, J. S., & Burton, R. R. Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science, 1978, 2, 155-192

Burstein, L. Analyzing multilevel educational data: The choice of an analytic model rather than a unit of analysis. In E. Baker, & E. Quellmalz (Eds.), Design, analysis, and policy in testing and evaluation. Beverly Hills, CA: Sage Publication, 1980a, 81-94.

Burstein, L. The role of levels of analysis in the specification of educational effects. In R. Dreeben & J. A. Thomas (Eds.), Analysis of educational productivity. Vol. 1: Issues in microanalysis. Cambridge, MA: Ballinger, 1980b, 119-190.

Burstein, L. Analysis of multilevel data in educational research and evaluation. In D. Berliner (Ed.), Review of Research in Education, Vol. 8, Washington, D.C.: American Educational Research Association, 1980c, 158-233.

Burstein, L. 1982

Burstein, L., & Miller, M. D. 1982

Burstein, L., Fischer, K., & Miller, M. D. The multilevel effects of background on science achievement at different levels of analysis: A cross-national comparison. Sociology of Education, 1980, 53(4), 215-255.

Burstein, L., Linn, R. L., & Capell, F. Analyzing multilevel data in the presence of heterogeneous within-class regressions. Journal of Educational Statistics, 1978, 3(4), 347-383.

Burstein, L., Miller, M. D., & Linn, R. L. The use of within-group slopes as indices of group outcomes. Unpublished manuscript, 1981.

California State Department of Education, 1976

Capell, F. Interpreting multilevel data from program evaluation. Paper presented at the Annual Meeting of the American Educational Research Asociation, San Francisco, CA, April 1979.

Capell, F. Interpreting multilevel data from program evaluations. Final Report, Santa Monica, CA: Rand Corporation, 1981.

Carver, R. P. Two dimensions of tests: Psychometric and edumetric. American Psychologist, 1974, 29, 512-518.

Carver, R. P. The Coleman Report: Using inappropriately designed tests. American Educational Research Journal, 1975, 12(1), 77-86.

Clark County School District. Elementary school management and accountability system handbook, Las Vegas, NV, 1976.

Cline, M. G., Ames, N., Anderson, R., Bale, R., Ferb, T., Joshi, M., Kane, M., Larson, J., Part, D., Proper, E., Stebbins, L., & Stern C. Education as experimentation: Evaluation of the Follow Through planned variation model, Vol. 1a, 1b. Cambridge, MA: Abt Associates, 1974.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, S., Weinfeld, F. D., York, R. L. Equality of educational opportunity (2 Vols.). Office of Education, U.S. Department of Health, Education and Welfare. Washington, D.C.: U.S. Government Printing Office, 1966.

Comber, L. C., & Keeves, J. P. Science education in nineteen countries, international studies in evaluation, Vol. 1. Stockholm: Almqvist and Wiksell; New York: Wiley, 1973.

Cooley, W. W., & Lohnes, P. R. Evaluation research in eduation: Theory principles and practices. New York: Irvington Publishers, Inc., 1976.

Cooley, W. W., Bond, L., & Mao, B. -J. Analyzing multilevel data. In R. A. Berk (Ed.), Educational evaluation methodology: The state of the art. Baltimore, MD: The Johns Hopkins University Press, 1981.

Corno, 1980

Cronbach, L. J. (with assistance of J. E. Deken & N. Webb). Research on classrooms and schools: Formulation of questions, designs and analysis. Occasional paper, Stanford Evaluation Consortium, Stanford, CA 1976.

Cronbach, L. J., & Snow, R. W. Aptitudes and instructional methods. New York: Irvington Press, 1977.

Cronbach, L. J., & Webb, N. Between-class and within-class effects in a reported aptitude x treatment interactions: Reanalysi of a study by G. L. Anderson. Journal of Educational Psychology, 1975 67, 717-724.

Dorr-Bremme, D. Assessing students: Teachers' routine practices and reasoning. In J. Burry, J. Catterall, B. Choppin, & D. Dorr-Bremme Testing in the nation's schools and districts: How much? What kinds: To what ends: At what costs? Report No. 194, Center for the Study of Evaluaton, University of California Los Angeles, 1982, 18-51.

Edmonds, R. R. Programs of school improvement: An overview. Paper presented at the National Invitational Conference on Research on Teaching: Implicatons for practice, Warrenton, VA, Februry 1982.

Felix, J. Supporting community efforts to improve instruction: A focus for evaluation in Cincinnati, unpublished manuscript, 1982.

Gastright, 1977.

Glaser, R. The future of testing: A research agenda for cognitive psychology and psychometrics. Technical Report No. 3. Learning Research and Development Center, Pittsburgh, PA, February 1981.

Goodlad, 1982

Greene, J. Individual and teacher/class effects in aptitude treatment studies. American Educational Research Journal, 1980, 17(3), 291-302.

Haney, W. Units of analysis issues in the evaluation of Project Follow Through. Unpublished report, Cambridge, MA: Huron Institute, 1974.

Haney, W. The Follow Through planned variation experiment, Vol. V. A Technical History of the National Follow Through Evaluation. Cambridge, MA: The Huron Institute, 1977.

Haney, W. Unit and levels of analysis in large-scale evaluation. In K. H. Roberts & L. Burstein (Eds.), Issues in aggregation no. & new directions for methodology of social and behavioral sciences, San Francisco: Jossey-Bass, Inc., 1980, 1-16.

Harnisch, D.L., & Linn, R.L. Analysis of item response patterns: Questionable test data and dissimilar curriculum paractices. Journal of Educational Measurement, 1981, 18(3), 133-146.

Kean, M. H., Summers, A. A., Raivetz, M. J., & Farber, I. J. What works in reading? Philadelphia, PA: School District of Philadelphia, 1979.

Klitgaard, R. W. 1974.

Klitgaard, R. W., & Hall, G. R. Journal of Human Resources, Vol. 10, Winter, 1975. (Also, A statistical search for unusually effective schools (R-1210-CC/RC). Santa Monica, CA: The Rand Corporation, 1973.)

Leinhardt, 1982.

Leinhardt, & Seewald, 1981.

Lohnes, P. Statistical descriptors of school classes. American Educational Research Journal, 1972, 9, 574-556.

Lyon, C.D., Doscher, L., McGranahan, P., & Williams, R. Evaluation and school districts. Center for the Study of Evaluation, University of California, Los Angeles, CA, December 1978.

Madaus, G. F., Airasian, P. W., & Kellaghan, P. School effectiveness: A reassessment of the evidence. New York: McGraw-Hill Book Company, 1980.

Madaus, Rakow & Kellaghan, 1980

Martin, J., Veldman, D. J., & Anderson, L. M. Within-class relationships between student achievement and teacher behaviors. American Educational Research Journal, 1980, 17(4) 479-490.

Merton, R. K., & Lazarsfeld, P. F. (Eds.) Continuities in social structure. Glenco, IL: Free Press, 1950.

Miller, M. D. Measuring between-group differences in instruction. Unpublished doctoral dissertation, University of California, Los Angeles, 1981.

Miller, M. D., & Burstein, L. 1979.

Monck (Munck???), 1981.

Peaker, G. F. An empirical study of education in twenty-one countries: A technical report. International Studies in Evaluation, Vol. 8. Stockholm, Sweden: Almqvist and Wiksell;, New York: Wiley, 1975.

Purkey & Smith, 1982

Rakow, E.A., Airasian, P.W., & Madaus, G.F. Assessing school and program effectiveness: Estimating teacher level effects. Journal of Eductional Measurement, 1978, 15(1), 15-22.

Ramey, M., Hillman, L., & Mathews, T. School characteristics associated with instructional effectiveness. Paper presented at the Annual Meeting of the American Educational Research Association, March 1982

Rasmussen, 1974

Rasmussen, 1976

Roberts, K.H., & Burstein, L. (Eds.) Issues in aggregation. No. 6, New directions for methodology of social and behavioral sciences. San Francisco, CA: Jossey-Bass, Inc., 1980.

Sato, T. 1978

Sato, T. Similarity of some indices of item response patterns (Research Memorandum E181-1). Tokyo: Computer and Communications System Research Laboratories, Nippon Electric Co., Ltd., 1981.

Schmidt, W. H. Examining the match between content tested and content taught. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA, April 1980.

Schmidt, 1983.

Seewald, 1981.

Shavelson, R. J. Teachers decision-making. In N.L. Gage (Ed.) The psychology of teaching methods, seventy-fifth yearbook of the National Society for the Study of Education, Part I, Chicago: University of Chicago Press, 1976, 372-414.

Sirotnik, K.A. Psychometric implications of the unit-of-analysis "problem" (with examples from the measurement of organizational climate). Journal of Educational Measurement, 1980, 17, 245-281.

Sirotnik, K.A., Nides, M.A., & Engstrom, B.A. Some methodological issues in developing measures of classroom learning environment. Studies in Educational Evaluation, 1980, 6, 279-289.

Sirotnik, K.A., & Oakes, J. A contextual appraisal system for schools: Medicine or madness? Educational Leadership, 1981a, 39, 164-173. (Also available as CSE Report).

Sirotnik, K.A., & Oakes, J. Toward a comprehensive educational appraisal system: A contextual perspective (Laboratory in School and Community Education Occasional Report No. 1). Los Angeles: University of California, 1981b.

Sirotnik, K. A., & Oakes, J. 1982

Spencer, B. D. On interpreting test scores as social indicators: Statistical considerations. Unpublished manuscript, 1982.

Stebbins, L. B., St. Pierre, R. G., Proper, E. C., Andersen, R. B., & Cerva, T. R. Education as experimentation: A planned variation model. Cambridge, MA: Abt Associates, 1977.

Swanson, 1976.

Tatsuoka, M. M. Recent psychometric developments in Japan: Engineers grapple with educational measurement problems. Paper presented at the ONR Contractors Meeting on Individualized Measurement, Columbia, MO, 1978.

Tatsuoka & Birenbaum, 1980

Thomas, J. A. Resource allocation in classrooms. National Institute of Education. Final Report No. G-74-0037, Washington, D.C., 1977.

Webb, N. M. Group process: The key to learning in groups. In K. H. Roberts and L. Burstein (Eds.), Issues in aggregation, Vol. 6, New directions for methodology of social and behavioral science, San Francisco: Jossey-Bass, 1980, 77-88.

Webb, N. M. 1980(b???a???)

Webb, N. M., & Shavelson, R. J. Multivariate generalizability of general educational development ratings. Journal of Educational Measurement, 1981, 18. 13-22.

Webb, N. M. 1982

White, et al, 1975

White, et al, 1976

Wiley, 1980

Wisenbaker & Schmidt, 1978 (1979???)